

Robust Kalman tracking and smoothing with propagating and non-propagating outliers

Peter Ruckdeschel

Fraunhofer ITWM, Abt. Finanzmathematik,
Fraunhofer-Platz 1, 67663 Kaiserslautern, Germany
and TU Kaiserslautern, AG Statistik, FB. Mathematik,
P.O.Box 3049, 67653 Kaiserslautern, Germany
e-mail: `peter.ruckdeschel@itwm.fraunhofer.de`

Bernhard Spangl

University of Natural Resources and Applied Life Sciences,
Institute of Applied Statistics and Computing,
Gregor-Mendel-Str. 33, 1180 Wien, Austria
e-mail: `bernhard.spangl@boku.ac.at`

Daria Pupashenko

Hochschule Furtwangen, Fak. Maschinenbau und Verfahrenstechnik,
Jakob-Kienzle-Strae 17, 78054 Villingen-Schwenningen, Germany
and TU Kaiserslautern, AG Finanzmathematik, FB. Mathematik,
P.O.Box 3049, 67653 Kaiserslautern, Germany
e-mail: `pud@hs-furtwangen.de`

March 2, 2013

Abstract

A common situation in filtering where classical Kalman filtering does not perform particularly well is tracking in the presence of propagating outliers. This calls for robustness understood in a distributional sense, i.e.; we enlarge the distribution assumptions made in the ideal model by suitable neighborhoods. Based on optimality results for distributional-robust Kalman filtering from Ruckdeschel (2001, 2010b), we propose new robust recursive filters and smoothers designed for this purpose as well as specialized versions for non-propagating outliers. We apply these procedures in the context of a GPS problem arising in the car industry. To better understand these filters, we study their behavior at stylized outlier patterns (for which they are not designed) and compare them to other approaches for the tracking problem. Finally, in a simulation study we discuss efficiency of our procedures in comparison to competitors.

KEYWORDS: 93E11, 62F35

1 Introduction

Motivation State space models (SSMs) build a flexible but still manageable class of dynamic models, and together with corresponding attractive procedures as the Kalman filter and its extensions they provide an immensely useful tool for a wide range of applications. In this paper we are focussing on an application in engineering in the context of a GPS problem arising in the car industry with different linear and non-linear state space formulations.

It is common knowledge for long that most of the classical procedures, used in this domain suffer from lack of robustness, in particular the Kalman filters and smoothers which we concentrate on.

The mere notion of robustness however in filtering context is not canonic, with the general idea to describe stability of the procedure w.r.t. variations of the “input parameters”. The choice of the “input parameters” to look at varies from notion to notion passing from initial values to bounds on user-defined controls to distributional assumptions. As general in robust statistics, we are concerned with distributional or outlier robustness; i.e.; our input parameters are the model distributions and defining suitable neighborhoods about the ideal model we allow for deviations in the respective assumptions which capture various types of outliers.

The amount of literature on robustifications of these procedures is huge, so we do not attempt to give a comprehensive account here. Instead, we refer to the surveys given in Ershov and Liptser (1978), Kassam and Poor (1985), Stockinger and Dutter (1987), Schick and Mitter (1994), Künsch (2001), and to some extent Ruckdeschel (2001, Sect. 1.5). In Section 3.2 below, we list references related more closely to our actual approach.

Problem statement In our context these outliers may be system-endogenous (i.e., propagating) or -exogenous (i.e. non-propagating), which induces the somewhat conflicting goals of tracking and attenuation.

If we head for robust optimality in the sense of minimizing mean squared error on neighborhoods, the standard outlier models from Fox (1972) pose problems barely tractable in closed form, although some approximate solutions have been given, e.g., in Masreliez and Martin (1977). A way out is given by outliers of substitutive type where closed-form saddle-points could be derived for the attenuation problem in filtering in Ruckdeschel (2001) which involve the hard-to-compute ideal conditional expectation. If however this conditional expectation is linear, we come up with an easy and fast robustification (rLS) which also is available in multi-variate settings. Although exact linearity can be disproved, approximate linearity usually holds, at least in a central region of the distribution.

This approach has been generalized to simple tracking problems where we can completely observe the states (i.e., the observation matrix is invertible) in Ruckdeschel (2010a,b). In this paper, we consider the general situation, i.e., tracking in situations where observation dimension is lower than the one of the states. To this end we propose a suitable generalization of the rLS to this tracking problem and discuss its limitations.

To be able to apply these techniques in the EM algorithm for parameter estimation in state space models (see Shumway and Stoffer (1982)) we also generalize our procedures to the fixed interval smoothing problem where we come up with a strictly recursive solution as in the classical case going backward from the last state to first one.

As needed in the application, we also make available our techniques for Extended Kalman filtering and smoothing, which to the best of our knowledge is novel.

Organization of the paper Our paper is organized as follows: In the text, the focus is on an understanding of the procedures and their behavior at data, so we delegate mathematical results

to the appendix. We put some emphasis though on complete specification of the procedures.

In Section 2, we introduce the general framework of linear, time varying SSMs and their extension as to outlier models. In Section 3 we define the filtering and smoothing problems and recall the classical optimal solutions, i.e., the Kalman filter and smoother and contrast these to our robustifications based on the rLS. This section also introduces the central procedure of this paper, the rLS.IO, and gives the necessary generalizations to cover the smoothing and the Extended Kalman filter case, and, for comparison, introduces other robust non-parametric alternatives. In Section 4, we study their behavior in the ideal situation and at stylized outlier patterns for which they are not necessarily designed. In particular we demonstrate that the tracking problem at a model with a non-trivial null space of the observation matrix leads to problems where, at least at filtering time, for certain outliers, any procedure must fail. Section 5 presents an application of our procedures in the car industry dealing with GPS problems arising there. We sketch three different SSMs used to model this situation, in particular including a non-linear one, making robust Extended Kalman filters indispensable. Section 6 provides a comparative simulation study in which the outliers are generated according to the ones for which the procedures have been defined. Conclusions are gathered in Section 7.

As to the mathematical details to our procedures, Appendix A.1 presents some derivation of the optimality of the Kalman filter among all linear filters without any rank-conditions for the arising matrices and under more general semi-norms than the Euclidean one. Appendix A.2 gives a brief account on the optimality of the rLS, citing relevant results. Finally, Appendix A.3 contains the mathematical core of this paper, giving the necessary generalizations in order to translate the optimality results from the AO case to the tracking problem.

2 Setup

2.1 Ideal model

We consider a linear SSM consisting in an unobservable p -dimensional state X_t evolving according to a possibly time-inhomogeneous vector autoregressive model of order 1 with innovations v_t and transition matrices $F_t \in \mathbb{R}^{p \times p}$, i.e.,

$$X_t = F_t X_{t-1} + v_t \quad (2.1)$$

a q -dimensional observation Y_t which is a linear transformation X_t involving an additional observation error ε_t and a corresponding observation matrix $Z_t \in \mathbb{R}^{q \times p}$,

$$Y_t = Z_t X_t + \varepsilon_t \quad (2.2)$$

In the ideal model we work in a Gaussian context, that is we assume

$$v_t \stackrel{\text{indep.}}{\sim} \mathcal{N}_p(0, Q_t), \quad \varepsilon_t \stackrel{\text{indep.}}{\sim} \mathcal{N}_q(0, V_t), \quad X_0 \sim \mathcal{N}_p(a_0, Q_0), \quad (2.3)$$

$$\{X_0, v_s, \varepsilon_t, s, t \in \mathbb{N}\} \text{ stochastically independent} \quad (2.4)$$

For this paper, we assume the hyper-parameters F_t, Z_t, Q_t, V_t, a_0 to be known.

2.2 Deviations from the ideal model

As announced, these ideal model assumptions for robustness considerations are extended by allowing (small) deviations, most prominently generated by outliers. In our notation, suffix “id” indicates the *ideal* setting, “di” the *distorting* (contaminating) situation, “re” the *realistic*, contaminated situation.

AOs and IOs In time series context the most important distinction to be made as to outliers is whether they propagate or not. To label these different types, for historical reasons, we use the terminology of Fox (1972) (albeit in a somewhat more general sense). Fox distinguishes *innovation outliers* (or IOs) which enter the state layer and hence propagate and *additive outliers* (or AOs) which only affect single observations and do not propagate. Originally, AOs and IOs denote gross errors affecting the observation errors and the innovations, respectively. We use these terms in a wider sense: *IOs* also may cover level shifts or linear trends which would not be included in the original definition. Similarly *AO* here will denote general exogenous outliers which do not propagate.

More specifically, our procedures rLS.AO and rLS.IO defined below both assume substitutive outliers, but should also provide protection to some extent to arbitrary outliers of (wide-sense) AO respectively IO type. To be precise, rLS.AO assumes a substitutive outlier (SO) model already used by Birmiwal and Shen (1993) and Birmiwal and Papantoni-Kazakos (1994), i.e.,

$$Y^{\text{re}} = (1 - U)Y^{\text{id}} + UY^{\text{di}}, \quad U \sim \text{Bin}(1, r) \quad (2.5)$$

for SO-contamination radius $0 \leq r \leq 1$ specifying the size of the corresponding neighborhood, and where U is assumed independent of (X, Y^{id}) and (X, Y^{di}) as well as

$$Y^{\text{di}}, X \text{ independent} \quad (2.6)$$

As usual, the contaminating distribution $\mathcal{L}(Y^{\text{di}})$ is arbitrary, unknown and uncontrollable.

Similarly rLS.IO assumes that (2.1) is split up into two steps,

$$\tilde{X}_t = F_t X_{t-1}^{\text{re}} + v_t^{\text{id}}, \quad X_t^{\text{re}} = (1 - \tilde{U}_t)\tilde{X}_t + \tilde{U}_t X_t^{\text{di}}, \quad Y_t^{\text{re}} = Z_t X_t^{\text{re}} + (1 - \tilde{U}_t)\varepsilon_t^{\text{id}} \quad (2.7)$$

where X_{t-1}^{re} is the state according to the contaminated past, v_t^{id} and $\varepsilon_t^{\text{id}}$ are an ideally distributed innovation respectively observation error, and \tilde{U}_t and X_t^{di} are defined in analogy to U_t and Y^{di} (i.e., with independence from all ideal distributions and the past).

Different and competing goals induced by AOs and IOs Due to their different nature, as a rule, a different reaction in the presence of IOs and AOs is required. As AOs are exogenous, we would like to ignore them as far as possible, damping their effect, while when there are IOs, something has happened in the system, so the usual goal will be to detect these structural changes as fast as possible.

A situation where both AOs and IOs may occur is more difficult, as we cannot distinguish IO from AO type immediately after a suspicious observation; it will be treated elsewhere.

Other deviation patterns Of course the two substitutive outlier types of Section 2.2 are by no means exhaustive: Trends and level shifts are not directly covered, neither are patterns where the outlier mechanism may know about the past like the patchy outliers described in Martin and Yohai (1986). Another type of outliers are outliers in the oscillation behavior, in both amplitude/scale and frequency, or more generally spectral outliers as arising in robustly estimating spectral density functions, compare Franke (1985); Franke and Poor (1984), and Spangl (2008). We try to account at least for some of them in Section 4.

3 Kalman filter and smoother and robust alternatives

Filter problem The most important problem in SSM formulation is the reconstruction of the unobservable states X_t by means of the observations Y_t . For abbreviation let us denote

$$Y_{1:t} = (Y_1, \dots, Y_t), \quad Y_{1:0} := \emptyset \quad (3.1)$$

Using MSE risk, the optimal reconstruction is the solution to

$$\mathbb{E} |X_t - f_t|^2 = \min_{f_t}, \quad f_t \text{ measurable w.r.t. } \sigma(Y_{1:s}) \quad (3.2)$$

We focus on filtering ($s = t$) in this paper, while $s < t$ makes for a prediction, and $s > t$ for a smoothing problem.

3.1 Classical method: Kalman filter and smoother

The general solution to (3.2), the corresponding conditional expectation $\mathbb{E}[X_t|Y_{1:s}]$ usually is rather expensive to compute. Hence as in the Gauss-Markov setting, restriction to linear filters is a common way out. In this context, Kalman (1960) introduced a recursive scheme to compute this optimal linear filter reproduced here for later reference:

$$\text{Initialization:} \quad X_{0|0} = a_0, \quad \Sigma_{0|0} = Q_0 \quad (3.3)$$

$$\text{Prediction:} \quad X_{t|t-1} = F_t X_{t-1|t-1}, \quad \Sigma_{t|t-1} = F_t \Sigma_{t-1|t-1} F_t^T + Q_t \quad (3.4)$$

$$\begin{aligned} \text{Correction:} \quad X_{t|t} &= X_{t|t-1} + K_t \Delta Y_t, & \Delta Y_t &= Y_t - Z_t x_{t|t-1}, \\ K_t &= \Sigma_{t|t-1} Z_t^T C_t^{-1}, & \Sigma_{t|t} &= (\mathbb{I}_p - K_t Z_t) \Sigma_{t|t-1}, \\ C_t &= Z_t \Sigma_{t|t-1} Z_t^T + V_t \end{aligned} \quad (3.5)$$

where $\Sigma_{t|t} = \text{Cov}(X_t - X_{t|t})$, $\Sigma_{t|t-1} = \text{Cov}(X_t - X_{t|t-1})$, and K_t is the so-called *Kalman gain*.

The Kalman filter has a clear-cut structure with an initialization, a prediction, and a correction step. Evaluation and interpretation is easy, as all steps are linear. The strict recursivity / Markovian structure of the state equation allows one to concentrate all information from the past useful for the future in $X_{t|t-1}$.

This linearity is also the reason for its non-robustness, as observations y enter unbounded into the correction step. A good robustification has to be bounded in the observations, otherwise preserving the advantages of the Kalman filter as far as possible.

3.2 A robustification of the least squares solution (rLS)

The idea of the procedures we discuss in this paper are based on *robustifying recursive Least Squares*: rLS, a filter originally introduced for AOs only, compare Ruckdeschel (2000, 2001). Starting from a different route of translating regression M estimators to SSM context, Boncelet and Dickinson (1983, 1987); Cipra and Romera (1991) arrive at similar weight functions, albeit with an iterative procedure to solve for the M equations. More recently, restricted to the particular SSM class of Holt-Winter-Forecasting, Gelper, Fried and Croux (2010) take up this idea and use regression M estimators, which, in addition to the present rLS approach involve recursive scale estimation. The closest recent approach stems from Cipra and Hanzak (2011), who—without translating their clipping to a bias side condition for an explicit outlier model—show a Lemma-5-type optimality (see Problem (A.8) below) for their procedure (with application to the special SSM case of exponential smoothing). Their solution is different from ours only in the choice of the norm used for clipping: They use a diagonal weighting matrix W whereas, in the AO case, we use no weighting; a weighting does appear though in the general IO solution, compare Lemma A.1 below.

In this paper, we also consider an IO-robust version of this concept, and hence here, in addition to the original definition, we append the suffixes “.AO” and “.IO” to distinguish the two versions. Let us begin with (wide-sense) AOs.

With only AOs, there is no need for robustification in the initialization and prediction step, as no (new) observations enter. As introduced in Ruckdeschel (2000), we robustify the

correction step, replacing $K\Delta Y$ by a Huberization $H_b(K\Delta Y)$ where $H_b(x) = x \min\{1, b/|x|\}$ for some suitably chosen clipping height b and some suitably chosen norm, natural candidates being Euclidean and Mahalanobis norm. It turns out that only little is gained if we account for this modification in the recursion for the filter covariance, so we leave this unchanged. That is, the only modification in the correction step becomes

$$X_{t|t} = X_{t|t-1} + H_b(K_t\Delta Y_t) \quad (3.6)$$

While this is a bounded substitute for the correction step in the classical Kalman filter, it still remains reasonably simple, is non iterative and hence especially useful for online-purposes.

However it should be noted that, departing from the Kalman filter and at the same time insisting on strict recursivity, we possibly exclude “better” non-recursive procedures. These procedures on the other hand would be much more expensive to compute.

As sketched in Appendix A.2, it can be shown that rLS not only is plausible, but also has some optimality properties. These are not the focus of this paper, though, and proofs of these properties appear elsewhere.

Choice of the clipping height b For the choice of b , we have two proposals. Both are based on the simplifying assumption that $E_{\text{id}}[\Delta X|\Delta Y]$ is linear, which in fact turns out to only be approximately correct. The first one chooses $b = b(\delta)$ according to an Anscombe (1960) criterion,

$$E_{\text{id}}|\Delta X - H_b(K\Delta Y)|^2 \stackrel{!}{=} (1 + \delta) E_{\text{id}}|\Delta X - K\Delta Y|^2 \quad (3.7)$$

where δ may be interpreted as “insurance premium” to be paid in terms of efficiency.

The second criterion uses the radius $r \in [0, 1]$ of the neighborhood $\mathcal{U}^{\text{so}}(r)$ (defined in (A.6)) and determines $b = b(r)$ such that

$$(1 - r) E_{\text{id}}(|K\Delta Y| - b)_+ \stackrel{!}{=} rb \quad (3.8)$$

This produces the minimax-MSE procedure for $\mathcal{U}^{\text{so}}(r)$ (compare Section A.2). Generalizing ideas of Rieder, Kohl and Ruckdeschel (2008), this criterion can be extended to situations where we only know that the radius lies in some interval, but we do not work this out here; for details see Ruckdeschel (2010b).

Remark 3.1. *It turns out that in filtering context, the second criterion reflects much better the problem inherent difficulty of a robustification: While 10% efficiency loss in the ideal model can be unreachable, because totally ignoring the new observation ΔY would only “cost” 5%, it may be much too little to produce a sizeable effect in other models. A radius of 0.1 however seems to lead to reasonable choices of b in most models.*

3.3 rLS.IO

As noted, in the presence of IOs, we want to follow an IO outlier as fast as possible.

Optimality results on distributional neighborhoods for the IO-case with the goal of a faster tracking to the best of our knowledge have not been considered by other authors so far.

The Kalman filter in this situation does not behave as bad as in the AO situation, but still tends to be too inert. To improve upon this, let us first simplify our model to the situation where we have an unobservable but interesting state $X \sim P^X(dx)$ and where instead of X we rather observe the sum

$$Y = X + \varepsilon \quad (3.9)$$

This equation reveals a useful symmetry of X and ε : Apparently

$$\mathbb{E}[X|Y] = Y - \mathbb{E}[\varepsilon|Y] \quad (3.10)$$

Hence we follow Y more closely if we damp estimation of ε , for which we use the rLS-filer. We should note that doing so, we rely on “clean”, i.e., ideally distributed errors ε . With the obvious replacements, our optimality results from the appendix translate word by word to a corresponding results for IOs, compare Ruckdeschel (2010a, Thm. 4.1).

In analogy to the definition of the rLS in equation (3.6), we set up an IO-robust version of the rLS as follows: We retain the initialization and prediction step of the classical Kalman filter and for the correction step, as proved in Corollary A.4, we note that in the ideal model, as $\mathbb{E}[\varepsilon_t|\Delta Y_t] = (\mathbb{I}_q - Z_t K_t)\Delta Y_t$, the correction step can also be written as

$$X_{t|t} = X_{t|t-1} + Z_t^\Sigma (\Delta Y_t - \mathbb{E}[\varepsilon_t|\Delta Y_t]), \quad (3.11)$$

where Z_t^Σ is a suitably generalized inverse for Z_t , minimizing the respective MSE among all $\mathbb{E}[\varepsilon|\Delta Y]$ -measurable (and square integrable) functions, i.e.,

$$Z_t^\Sigma := \Sigma_{t|t-1} Z_t^\top (Z_t^\top \Sigma_{t|t-1} Z_t)^- \quad (3.12)$$

The IO robustification then simply consists in replacing $\mathbb{E}[\varepsilon_t|\Delta Y_t]$ by $H_b((\mathbb{I}_q - Z_t K_t)\Delta Y_t)$, i.e., the rLS.IO correction step as

$$X_{t|t} = X_{t|t-1} + Z_t^\Sigma [\Delta Y_t - H_b((\mathbb{I}_q - Z_t K_t)\Delta Y_t)] \quad (3.13)$$

Here the same arguments for the choice of the norm and the clipping height apply as for the AO-robust version of the rLS.

To better distinguish IO- and AO-robust filters, let us call the IO-robust version *rLS.IO* and (for distinction) the AO-robust filter *rLS.AO* in the sequel.

Remark 3.2. *It is worth noting that also our IO-robust version is a filter, hence does not use information of observations made after the state to reconstruct; rLS.IO is strictly recursive and non iterative, hence well-suited for online applications.*

3.4 Extended Kalman filter

In the application we are heading for, in Model (M3) below, the SSM given by equations (2.1) and (2.2) is only a linearization of a nonlinear SSM given by

$$X_t = f_t(X_{t-1}, u_t, v_t) \quad (3.14)$$

$$Y_t = z_t(X_t, w_t, \varepsilon_t) \quad (3.15)$$

for smooth, known functions f_t and z_t in the states X_t , in the innovations v_t , in the observation errors ε_t and some user defined controls u_t and w_t . With $\bar{v}_t = \mathbb{E} v_t$, $\bar{\varepsilon}_t = \mathbb{E} \varepsilon_t$, this is linearized to give the following *Extended Kalman filter* taken from Wan and van der Merwe (2002)

$$\text{Initialization: } X_{0|0} = a_0, \quad \Sigma_{0|0} = Q_0 \quad (3.16)$$

$$\text{Prediction: } X_{t|t-1} = f_t(X_{t-1|t-1}, u_t, \bar{v}_t), \quad \Sigma_{t|t-1} = F_t \Sigma_{t-1|t-1} F_t^\top + B_t Q_t B_t^\top \quad (3.17)$$

$$\text{for } F_t = \frac{\partial}{\partial x} f_t(x, u_t, \bar{v}_t)|_{x_{t-1|t-1}}, \quad B_t = \frac{\partial}{\partial v} f_t(X_{t-1|t-1}, u_t, v)|_{\bar{v}_t}$$

$$\begin{aligned} \text{Correction: } X_{t|t} &= X_{t|t-1} + K_t \Delta Y_t, & \Delta Y_t &= Y_t - z_t(X_{t|t-1}, w_t, \bar{\varepsilon}_t), \\ K_t &= \Sigma_{t|t-1} Z_t^\top C_t^{-1}, & \Sigma_{t|t} &= (\mathbb{I}_p - K_t Z_t) \Sigma_{t|t-1}, \\ C_t &= Z_t \Sigma_{t|t-1} Z_t^\top + D_t V_t D_t^\top, & & \end{aligned} \quad (3.18)$$

$$\text{for } Z_t = \frac{\partial}{\partial x} z_t(x, u_t, \bar{v}_t)|_{x_{t|t-1}}, \quad D_t = \frac{\partial}{\partial \varepsilon} z_t(X_{t|t-1}, u_t, \varepsilon)|_{\bar{\varepsilon}_t}$$

As to the robustification of this algorithm by rLS.AO and rLS.IO, as in the linear case, we simply replace the term $K_t \Delta Y_t$ by $H_b(K_t \Delta Y_t)$ in the AO case and by $Z_t^\Sigma (\Delta Y_t - H_b(\mathbb{I}_q - Z_t K_t) \Delta Y_t)$ in the IO case.

3.5 A robust smoother

In many situations, in particular for an application of the EM-algorithm to estimate the hyperparameters, it is common use to enhance the filtered values $X_{t|t}$ in retrospective, accounting for the information of $Y_{t+1:T}$ which is available in the mean time. To retain recursivity, we use a corresponding backward recursion as to be found in Anderson and Moore (1990, Sec.7.4, (4.5)):

$$X_{t|T} = X_{t|t} + J_t(X_{t+1|T} - X_{t+1|t}), \quad J_t = \Sigma_{t|t} F_t^\tau \Sigma_{t+1|t}^{-1} \quad (3.19)$$

with respective update for the smoothing covariance,

$$\Sigma_{t|T} = \Sigma_{t|t} + J_t(\Sigma_{t+1|T} - \Sigma_{t+1|t})J_t^\tau \quad (3.20)$$

Writing (3.19) as $X_{t|T} - X_{t|t} = J_t[(X_{t+1|T} - X_{t+1|t+1}) + (X_{t+1|t+1} - X_{t+1|t})]$, we see that the first summand in the brackets of the RHS is just the preceding iterate of the LHS in the recursion and the second summand is just the already robustified increment of the correction step in the filter. Hence, sticking to our outlier models for IO and AO contamination which independently affect single X_t 's and Y_t 's the modification only has to be done in the (new) treatment of ΔY_t , i.e., in the second summand, so there is no further need for robustification in the backwards loop.

3.6 A nonparametric approach: median based filters

A nonparametric approach to assess the robustness issues met with Kalman filtering, which does not use any state space formulation is to use median-type filters for this purpose. Standard median filters remove outliers and preserve level shifts but their output does not properly represent linear trends. The repeated median (RM) has been developed for the extraction of monotonic trends with intercept and slope from a time series.

In order to combine advantages of several specifications of location and regression based median-type filters, a variety of hybrid filters has been introduced in Fried, Bernholt and Gather (2006). From these filters, in Section 6, we use the regression based and predictive approach PRMH. In addition, we use a location based approach MMH for smoothing (notation MMH taken over from R package `robfilter`, Fried and Schettlinger (2010)).

Because of their high breakdown points for window width k of $(\lfloor k/2 \rfloor + 1)/n$ for PRMH and $(2\lfloor k/2 \rfloor + 1)/n$ for MMH, see Fried, Bernholt and Gather (2006), these hybrid RM-type filters and smoothers are prominent candidates for a first sweep over the data.

For an automatic choice of the window width, the Adaptive On-line Repeated Median Regression (ADORE) filter (Schettlinger, 2009) provides robust on-line extraction by a moving window technique with adaptive window width selection.

In our setting, the median-based filters as they stand can only be applied to situations with one-dimensional observations; generalizations to higher dimensions beyond mere coordinate-wise treatment have been introduced in Schettlinger (2009), but are not dealt with in this paper. Being non-parametric, these filters know nothing about an underlying SSM, so cannot be used for state reconstruction, except for the case that $Z_t = 1$ (or for $p = q$, \mathbb{I}_q in higher dimensions). So we include these filters only in the steady state examples of Sections 4.2, 4.3 and 6 (Model (SimB)).

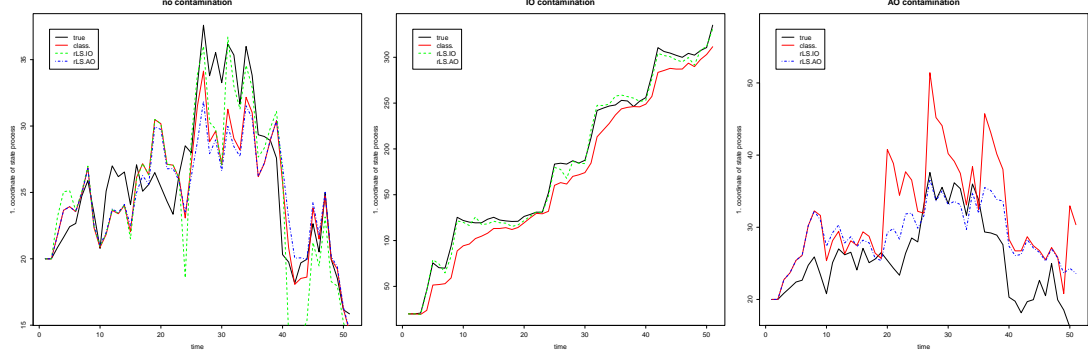


Figure 1: Typical results of the state-space model for different situations: left no contamination, middle IO contamination, right AO contamination

4 Behavior of the filters at stylized outlier situations

In this section, we study the behavior of our filters in the ideal situation and at stylized outlier patterns for which they are not necessarily designed—these cover AOs and IOs in the original sense, their behavior at trends and level shifts and at changes in the oscillation behavior. We also look at the effects of a non-trivial null space of the observation matrix such that the information given by the observation does not suffice to reconstruct the state.

4.1 The ideal situation, spiky outliers and IOs

To start with, we study the behavior of the robustified (extended) Kalman filter (EKF) in three different situations, namely no contamination, contamination by (classical) IOs and AOs, and use the following hyper parameters for the SSM:

$$\begin{aligned} a_0 &= \begin{pmatrix} 20 \\ 0 \end{pmatrix}, & Q_0 &= \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \\ F_t &= \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}, & Z_t &= \begin{pmatrix} 0.3 & 1 \\ -0.3 & 1 \end{pmatrix}, & Q_t &= \begin{pmatrix} 0 & 0 \\ 0 & 9 \end{pmatrix}, & V_t &= \begin{pmatrix} 9 & 0 \\ 0 & 9 \end{pmatrix}. \end{aligned} \quad (4.1)$$

We note that the first coordinate of the above state process is a random walk and therefore non-stationary, whereas the second coordinate is just white noise.

The innovations v_t and the errors ε_t of the observation process are simulated from a contaminated bivariate normal distribution

$$\mathcal{CN}_2(r, 0, R, \mu_c, R_c) = (1 - r)\mathcal{N}_2(0, R) + r\mathcal{N}_2(\mu_c, R_c), \quad (4.2)$$

with r , the amount of contamination, set to 10% and where $R = Q_t$ in case of the innovations v_t and $R = V_t$ in case of the observation errors ε_t , and the moments of the contaminating distribution are given by

$$\mu_c^T = (25, 30), \quad R_c = \text{diag}(0.9, 0.9). \quad (4.3)$$

Then the bivariate filter estimates of the three simulated state-space models are computed using the classical Kalman and rLS filters. In Figure 1, we show typical realizations of the state process together with their filter estimates for different contamination situations. Only the first

coordinate of the estimated state vector is plotted. The thick black line is the true state process, while classical Kalman, rLS.IO, and rLS.AO filters are plotted as light red line, dotted green line, and dot-dashed blue line, respectively. We note that contamination by additive outliers cannot be seen directly in Figure 1 because only the observation equation is affected. However, the spikes, especially visible for the filter estimate of the classical Kalman filter, do indicate additive outliers.

Ideal situation: The classical Kalman filter as well as its robustified versions, yield almost the same results.

IO contamination: The rLS.IO filter is able to follow the true state almost immediately, whereas the classical Kalman filter is only able to track the true state with a certain delay.

Spiky outliers: The rLS.AO filter is not affected by additive outliers whereas the classical Kalman filter is prone to them.

4.2 Changes in oscillation pattern and level shifts

Next, in situations with outliers scaling on different frequencies, we compare our robustified versions of the Kalman filter to a non-parametric filtering method, i.e., the ADORE filter proposed by Schettlinger (2009)—with an automatic selection of the window width. As outlier situations, we consider (classical) IOs and AOs (with contamination radius $r = 10\%$ each) as well as the special endogenous case where we substitute part of the state by a completely artificial signal. For the SSM we use an AR(2) process, i.e., an autoregressive process of order 2, as state process, and the following hyper parameters:

$$\begin{aligned} a_0 &= \begin{pmatrix} 0 \\ 0 \end{pmatrix}, & Q_0 &= \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \\ F_t &= \begin{pmatrix} 1 & -0.9 \\ 1 & 0 \end{pmatrix}, & Z_t &= \begin{pmatrix} 1 & 0 \end{pmatrix}, & Q_t &= \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, & V_t &= \begin{pmatrix} 1 \end{pmatrix}. \end{aligned} \quad (4.4)$$

The innovations v_t in the IO situation again stem from a contaminated bivariate normal distribution given in (4.2) with moments of the contaminating distribution given by

$$\mu_c^\tau = (30, 0), \quad R_c = \begin{pmatrix} 0.1 & 0 \\ 0 & 0 \end{pmatrix}. \quad (4.5)$$

The contaminating distribution of the errors ε_t of the observation process is $\mathcal{N}(10, 0.1)$. Then, for all different situations the filter estimates are computed using the rLS filter and the ADORE filter. Moreover, also the classical Kalman filter estimates were calculated.

In Figure 2, the realizations of the state process, together with their filter estimates are displayed for different situations of contamination. The black line is again the true state process, while classical Kalman, rLS, and ADORE filters are plotted by a red line, a dashed green line, and a dot-dashed blue line, respectively.

Endogenous contamination by artificial signal: In this case, whole parts of the state process are substituted by an artificial signal, more specifically by a so called block signal (cf., e.g., Donoho and Johnstone, 1994) which is piecewise constant with random length and amplitude. Here, the rLS.IO filter is able to track the state process best. The classical Kalman filter, as already seen in the previous section, is not able to follow the level shifts. Using the ADORE filter as non-parametric filtering method an obvious time delay in following the state signal can be seen.

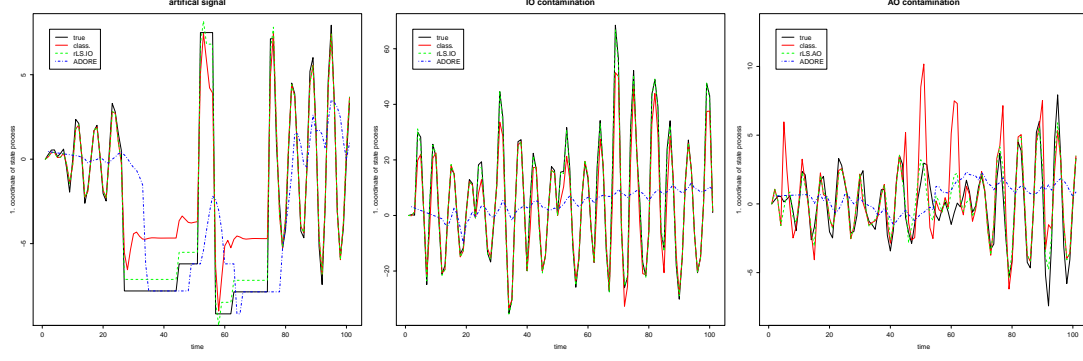


Figure 2: Results of the simulated state-space model for different situations: left contamination by artificial signal, middle IO contamination, right AO contamination

IO contamination: Here, the rLS.IO filter is able to follow the true state, whereas the classical Kalman filter fails to track the spikes of the state signal. The ADORF filter as non-parametric alternative only fits an overall trend. In general, non-parametric filters as ADORF are specialized to reveal trends, trend changes or shifts of an underlying, possibly non-stationary signal in the presence of outliers and, according to our experience, they extremely smooth the underlying process.

Spiky outliers: The rLS.AO filter is not affected by AOs whereas the classical Kalman filter is prone to them. The ADORF filter again shows the above mentioned properties and estimates more or less an overall trend of the underlying AR(2) process.

4.3 Coping with non observed aspects

One important aspect of reconstructing states in SSMs is observability, compare Anderson and Moore (1990, App. C), as in general matrix Z_t will have a non-trivial null space with the consequence that state signals falling to this null space are not visible at filtering time. Smoothing may to some extent relieve this problem with a certain time delay, when subsequent transitions F_s move the states in such a way that they become visible to Z_s at a later stage.

We illustrate this problem studying how the considered versions of the Kalman filter cope with such non observed aspects in the following setup:

$$T = 50, \quad F = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \quad Z = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad Q = \text{diag}(0, 0, 0.001), \quad (4.6)$$

$$V = \text{diag}(0.1, 0.001), \quad a_0 = (0, 0, 0)', \quad Q_0 = \text{diag}(1, 0.1, 0.001).$$

Here, we use as outlier specification the ones from Section 2.2 with $r_{\text{IO}} = r_{\text{AO}} = 0.1$, i.e., (2.5) and (2.6) in the AO case and (2.7) in the IO case. As contaminating distributions we chose $X_t^{\text{di}} \sim \text{multiv.Cauchy}(0, Q)$ and $Y_t^{\text{di}} \sim \text{Cauchy}/1000$ (using R packages `mvtnorm` (Genz et al., 2011) and `MASS` (Venables and Ripley, 2002)).

In Figure 3, for different versions of the rLS, we display one realization of the state process in the ideal and in the real IO-contaminated situation together with the filter and smoother estimates. The black line is the true state process in the ideal situation, while the red line represents the IO-contaminated state process, i.e., the real situation. The dashed green line represents the rLS filter, the dot-dashed blue line the corresponding rLS smoother.

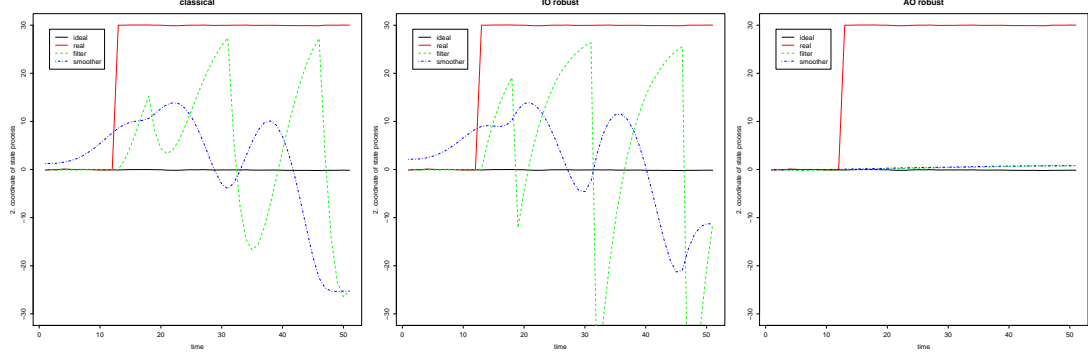


Figure 3: Filter estimates of the simulated state-space model using different filter and smoothers: left classical Kalman filter and smoother, middle IO-robust filter, right AO-robust filter

Only the second coordinate, which lies in $\ker Z_t$, of the state process is plotted.

All three plots clearly reveal that none of the proposed filters is able to cope with this situation, i.e., to correctly follow the level shift in the second coordinate of the state process caused by an IO-contamination.

5 Application

In this section, we show an application of the procedures described above to data from a cooperation with the department for Mathematical Methods in Dynamics and Durability at Fraunhofer ITWM and Kaiserslautern University, in particular with Nikolaus Ruf and Jürgen Franke. This data is part of a larger project they are involved. In the application, we deal with data taken from a vehicle moving on some track, which consists of four data channels, i.e.,

- time recorded in seconds ($[s]$);
- vehicle speed \widetilde{sp}_t , measured in meters per second ($[m/s]$);
- measured altitude \widetilde{h}_t ($[m]$);
- pitch angle speed $\widetilde{\alpha}_t$, measured in radians per second ($[rad/s]$).

We are interested in slope estimation, more precisely in the change of altitude over distance, because this information cannot be captured so easily from cartographic information.

The original data has been distorted by different factors, e.g. GPS measurement error, discreteness of the measurement process, change of the road surface, etc. If during the measurement the process signal was lost for short time because of tunnels on the way or for some other reasons, some part of the data is missing and we can observe e.g. sudden jumps of the altitude or speed, what is impossible in practice for the tracks. Such situations are very common in reality, therefore we are interested in constructing reliable parametric models of such errors in order to reconstruct the data.

With different simplifying assumptions on the observed data—to some degree for illustration purpose—we construct three models:

(M1) **A linear time-invariant model** with the following state vector and matrix:

$$X_t = \begin{pmatrix} h_t \\ \dot{h}_t \\ c_t \end{pmatrix} \quad F = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix} \quad v_t \sim N_3(0, Q), \quad Q = \text{diag}(0, 0, 0.01),$$

where $c_t = sp_{t+1}\dot{\alpha}_t, t = 0, \dots, T-1$ denotes the compound increment measured in $([m/s] \times [rad/s])$. Note that by construction, c_T is not defined, therefore we additionally assume that $c_T = sp_T\dot{\alpha}_T$.

The hyper-parameters of the observation equation in this model are the following

$$Y_t = \begin{pmatrix} \tilde{h}_t \\ \tilde{c}_t \end{pmatrix} \quad Z = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \varepsilon_t \sim N_2(0, V), \quad V = \text{diag}(5, 0.01).$$

In reality, the variances have been obtained by an application of an EM-type algorithm as the one by Shumway and Stoffer (1982). We mention that this EM-Algorithm is all but robust, being based on classical estimators for first and second moments. This robustness issue will be discussed elsewhere in more detail.

In our application, though, we inspected the norms of the observation residuals ΔY_t and the estimated innovations \hat{v}_t and errors $\hat{\varepsilon}_t$ which were obtained from the rLS-filters and -smoothers applied to the real data. Among these obtained $\|\Delta Y_t\|$, $\|\hat{v}_t\|$, $\|\hat{\varepsilon}_t\|$, we did not observe outstanding values, which indicates that application of the classical EM-Algorithm in this case should be justified.

The initial distribution of the state vector here is defined as follows

$$a_0 = (h_1, 0, 0)', \quad Q_0 = \text{diag}(5, 1, 0.01).$$

In Figure 4, we plot the actual observations together with the respective reconstruction by means of the classical Kalman filter, the rLS.AO, and the rLS.IO in Models (M1) and (M2) defined below, with a zoom-out of observations 100–130 to better distinguish the different models. The classical reconstruction is not clearly visible at the plot, since in case of these data, it almost coincides with the rLS.AO.

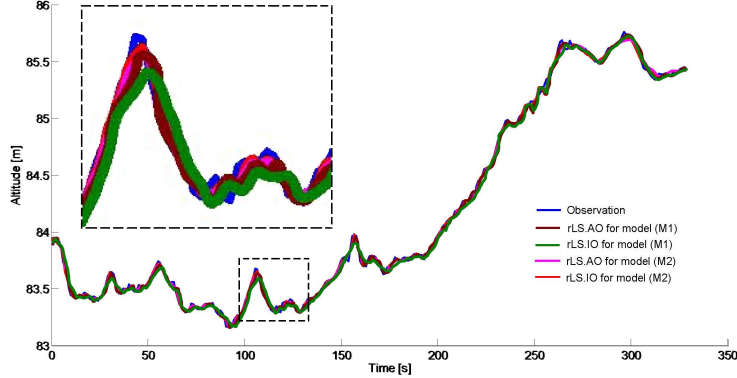


Figure 4: Observations and reconstructed values for different filters at Models (M1) and (M2)

Other data sets we analyzed showed much stronger evidence for outliers, but for confidentiality reasons, we limit ourselves to this data set and instead discuss the behavior of our

procedures at outliers in generated data. We may mention though, that our procedures work well in the discussed problems.

(M2) **A linear time-varying model** with the following state hyper-parameters:

$$X_t = \begin{pmatrix} h_t \\ \alpha_t \\ \dot{\alpha}_t \end{pmatrix} \quad F_t = \begin{pmatrix} 1 & sp_t \Delta t & 0 \\ 0 & 1 & \Delta t \\ 0 & 0 & 0 \end{pmatrix} \quad v_t \sim N_3(0, Q), \quad Q = \text{diag}(0, 0, 0.05).$$

Note that to complete the derivation of the model, we need to define a state matrix for $t = 0$, and since we do not have the vehicle speed at time $t = 0$ we set it to the first observation sp_1 , i.e.

$$F_0 = \begin{pmatrix} 1 & sp_1 \Delta t & 0 \\ 0 & 1 & \Delta t \\ 0 & 0 & 0 \end{pmatrix}.$$

The hyper-parameters of the observation equation in this model are the following:

$$Y_k = \begin{pmatrix} \tilde{h}_k \\ \tilde{\alpha}_k \end{pmatrix} \quad Z = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \varepsilon_t \sim N_2(0, V), \quad V = \text{diag}(5, 0.005).$$

The initial distribution of the state vector is again defined as follows

$$a_0 = (h_1, 0, 0)', \quad Q_0 = \text{diag}(5, 0.005, 0.005).$$

(M3) **A quadratic time-invariant model** accounting for acceleration. This gives a nonlinear SSM in the notation of Section 3.4, specified as

$$f_t(X_{t-1}, u_t, v_t) = A(X_{t-1} \otimes X_{t-1}) + BX_{t-1} + v_t,$$

$$z_t(X_t, w_t, \varepsilon_t) = ZX_t + \varepsilon_t$$

Here \otimes denotes the Kronecker product, i.e., the quadratic term can be written in the following form

$$A(X_t \otimes X_t) = \sum_{l=1}^p A^l [X_t]_l X_t, \quad k = 0, \dots, T-1$$

with known $p \times p^2$ -matrix $A = (A^1 \mid \dots \mid A^p)$.

The hyper-parameters, states and observations of constructed for this application quadratic time-invariant model are the following

$$X_t = \begin{pmatrix} h_t \\ sp_t \\ \dot{sp}_t \\ \alpha_t \\ \dot{\alpha}_t \end{pmatrix} \quad A = \begin{pmatrix} 0 & 0 & 0 & \Delta t & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0_{5 \times 5} \mid 0 & 0 & 0 & 0 & 0 \mid 0_{5 \times 5} \mid 0_{5 \times 5} \mid 0_{5 \times 5} \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

$$B = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & \Delta t & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & \Delta t \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \quad v_t \sim N_5(0, Q), \quad Q = \text{diag}(0, 0, 2, 0, 0.005),$$

$$Y_t = \begin{pmatrix} \tilde{h}_t \\ \tilde{sp}_t \\ \tilde{sp}_t \\ \tilde{\alpha}_t \end{pmatrix} \quad Z = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \quad \varepsilon_t \sim N_4(0, V), \quad V = \text{diag}(5, 2, 2, 0.005).$$

The initial distribution of the state vector here is defined as follows

$$a_0 = (h_1, sp_1, 0, 0, 0)', \quad Q_0 = \text{diag}(5, 2, 2, 0.005, 0.005).$$

6 Simulation

To see how our procedures (and some competitors) work in the outlier setting they are constructed for, we produce an simulation study done in R, R Development Core Team (2012), in the framework of our package `robKalman` developed at <http://r-forge.r-project.org/projects/robkalman/>.

For these simulations, we generated 10000 runs of data for two different models: Model (SimA), a one-dimensional time-invariant steady-state model with parameters $F = Z = Q = V = 1$, $a_0 = 1$ and time horizon $T = 50$ for comparison with the median-based filters of Section 3.6. More specifically, we use function `hybrid.filter` from R package `robfilter`, with specifications PRMH and MMH denoted by `hybf` and `hybs` below, respectively (for *hybrid filter / smoother*). Both procedures are used with fixed window width 5 and minimum number of non-missing observations 2, reflecting the actual outlier situation.

Second, we study Model (SimB) which takes over dimensions and typical parameters from the application of Section 5 as specified in detail in (4.6) and already used in Section 4.3.

As outlier specification, in both models we use the ones from Section 2.2 with $r_{\text{IO}} = r_{\text{AO}} = 0.1$, i.e., (2.5) and (2.6) in the AO case and (2.7) in the IO case. As contaminating distributions we chose Cauchy for $X_t^{\text{di}} \sim \text{Cauchy}(-10, 1)$ and $Y_t^{\text{di}} \sim \text{Cauchy}(5, 1)$ in Model (SimA), while in Model (SimB) we used $X_t^{\text{di}} \sim \text{multiv.Cauchy}(0, Q)$ and $Y_t^{\text{di}} \sim \text{Cauchy}(0, 1/1000)$ (using R packages `mvtnorm` (Genz et al., 2011) and `MASS` (Venables and Ripley, 2002)).

The results for Model (SimA) are summarized in Table 1 and Figures 5–8, the ones for Model (SimB) in Table 2 and Figures 9–12. In the annotation we denote the smoother versions of rLS.AO and rLS.IO by SrLS.AO and SrLS.IO, respectively.

In the tables, we display the empirical mean squared error (MSE) for each of the procedures at each situation. It is clearly visible that each filter does the job well it is made for: The classical Kalman filter is best in the ideal situation with rLS.IO and (a little less so) rLS.AO still close by. In the AO situation, the AO-robust filter is best (also compared to the median-type filters which comes second best), whereas classical Kalman, and, even worse rLS.IO, have problems. In the IO situation the situation changes dramatically: rLS.IO excels, whereas the classical Kalman filter shows its inertia, and, much worse, rLS.AO and `hybf` are not at all able to track these abrupt signal changes. For smoothing the situation is a bit different: At large we have the same picture as for filtering, however the IO-robust smoother is not doing a good job at all: the filter is much better here. This remains to be studied in more detail. In the ideal and AO situation however the smoothers do improve the filter (as they should). So it seems a strictly recursive smoother like the one we propose is well capable to deal with spiky outliers but much less so to track abrupt changes.

As to the graphics, in Figures 5 and 9, we display the coordinatewise distribution of the reconstruction errors $\Delta X_{35} = X_{35|35} - X_{35}$ for filtering and $\Delta X_{35} = X_{35|50} - X_{35}$ for smoothing.

The columns of the panels are the situations (ideal, AO, and IO), the rows (for Model (SimB)) the state coordinates. In each panel, we display the filters first and then the smoothers, ordered as in the columns of the tables. In order to keep the boxes of the boxplots distinguishable, we skipped all outliers outside $(-15, 15)$ in Model (SimA) and 8 times the coordinate-wise maximal interquartile range in Model (SimB). In addition to the tables of the MSE, these figures reveal that also in terms of the bulk of the data there are differences among the procedures: In Figure 9 for Model (SimB), coordinate 1 is hardest to reconstruct, and shows large differences between smoother and filter. At least for coordinates 1 and 2, the central boxes are definitively smallest for the rLS.AO filter and its smoother in the AO situation, with even a large improvement by the smoother. The same goes for the rLS.IO filter in the IO situation, whereas as already noted in the tables, the IO smoother is less convincing in the IO situation, although not really worse than the filter in terms of the central box, although coordinate 3 tells a different story here.

In Figure 5 for Model (SimA), we scaled all panels to the same coordinates in order to be better able to compare the situations; otherwise the conclusions to be drawn parallel the ones for Model (SimA), except that in positions 4 and 8 in each panel we display the median type filters and smoother which lead to largest central boxes in the filters, and in the smoothers are in between the specialized robust smoother and the respective unsuitably robustified smoother.

The remaining figures display the distribution of the normed reconstruction errors, where in spite of Proposition A.5 below, we have not changed the norm in the IO situation. To visualize all outliers this time, we use a logarithmic scale for the y -axis which also emphasizes “inliers” (compare Hampel et al. (1986, p. 140)), i.e., situations where the procedures behave extraordinarily well. Per se, for reconstruction, this is of course a beneficiary situation, but, in case inference is also of interest, this will lead to underestimation of the true variation.

In Model (SimA) in the ideal and IO situation, the median-type filters and smoothers excel in this direction, but also the rLS filters both perform better than the classical filter in this respect. For the smoother, the classical one is best, though. To the upper tail, the norms even for the worst situations surpasses the central box only by little (on log-scale, though). In the AO situation the specialized robust filters and smoothers have the shortest tail but as to the boxes are hardly distinguishable from the classical ones, while in the IO situation only the rLS.IO has a convincing tail.

In Model (SimB), in the ideal situation, the smoothers are clearly better than the filters but much less so in the outlier situations (although the AO smoother is significantly better than the filter). We also see that the classical procedures produce more inliers than the robustified ones in the ideal and IO situation, while in the AO situation no clear statement can be made.

situation	filter				smoother			
	Kalman	rLS.IO	rLS.AO	hybf	Kalman	rLS.IO	rLS.AO	hybs
id	0.628	0.679	0.808	2.180	0.447	0.475	0.575	0.939
AO	24.206	30.379	1.446	4.950	14.087	14.226	1.097	1.475
IO	30.175	0.729	1850.977	629.129	66.538	95.579	1846.870	623.371

Table 1: empirical MSEs in Model (SimA) at $t = 35$ for $T = 50$

7 Discussion and conclusion

Contribution In this paper we have presented robustifications of the Kalman filter and smoothers specialized either for damping of spiky outliers or for faster tracking a deviated series where the smoothers and the general IO-robust filter are new to the best of our knowledge.

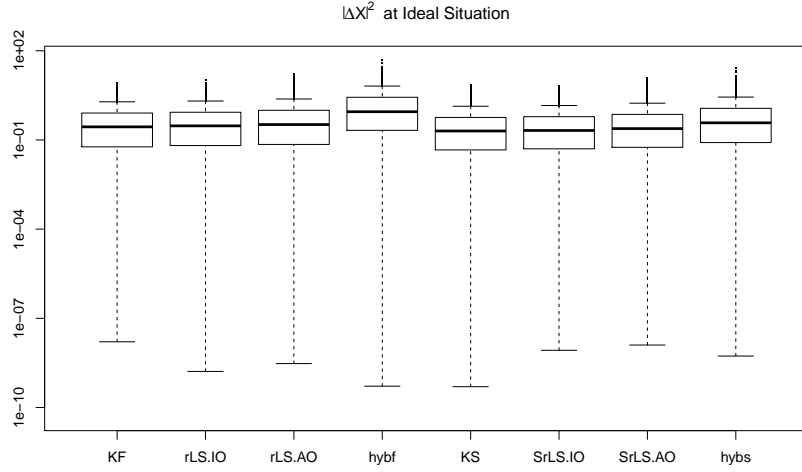


Figure 5: $\|\Delta X_t\|^2$ in Model (SimA) at $t = 35$ for $T = 50$ in ideal situation; filtered (boxes 1–4) and smoothed ($T = 50$; boxes 5–8) versions

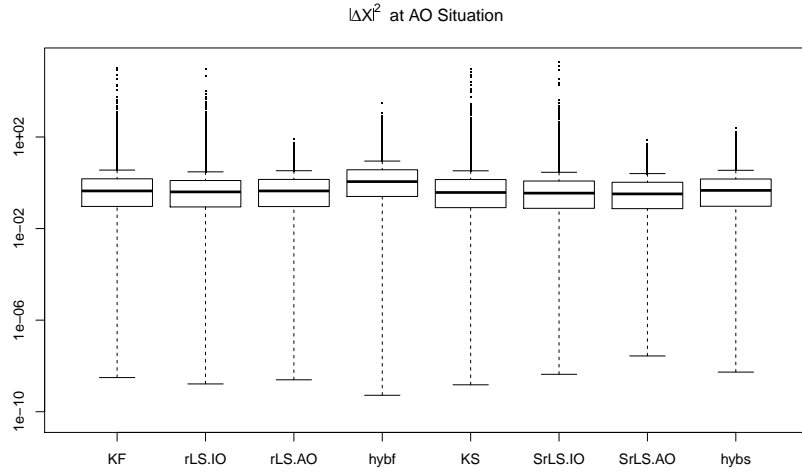


Figure 6: $\|\Delta X_t\|^2$ in Model (SimA) at $t = 35$ for $T = 50$ in AO situation; filtered (boxes 1–4) and smoothed ($T = 50$; boxes 5–8) versions

All our procedures are recursive, hence extra-ordinarily fast, so can be used in online problems and can easily be used to also robustify the Extended Kalman filter for non-linear state-space models. Our procedures are implemented to R, developed under **r-forge**, <http://r-forge.r-project.org/> and soon will be submitted to CRAN, <http://cran.r-project.org/>.

We have demonstrated the superior behavior of our procedures at the outlier situations they are made for and, in a study of stylized “realistic” outlier situations we showed that they

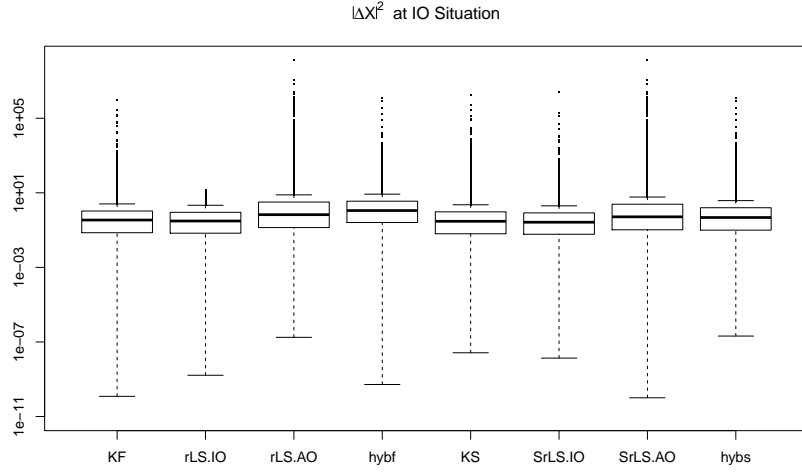


Figure 7: $\|\Delta X_t\|^2$ in Model (SimA) at $t = 35$ for $T = 50$ in IO situation; filtered (boxes 1–4) and smoothed ($T = 50$; boxes 5–8) versions

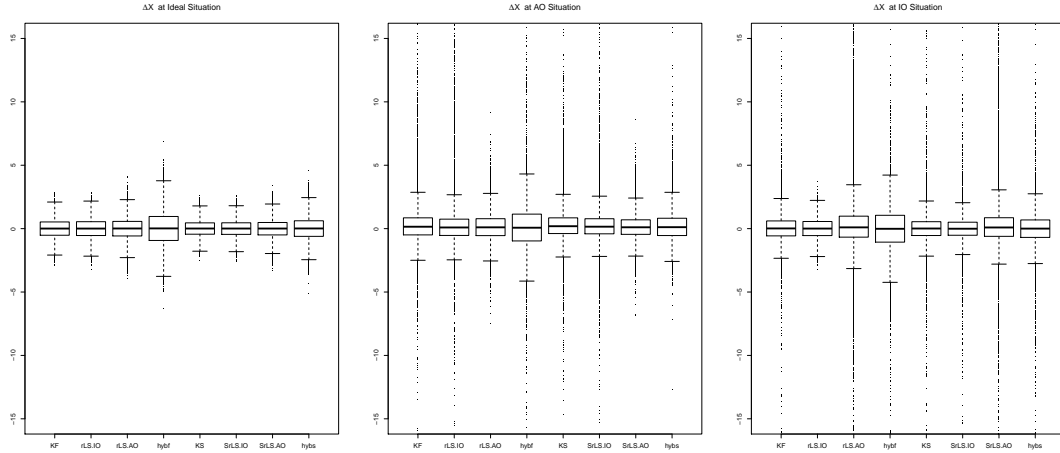


Figure 8: ΔX_t in Model (SimA) at $t = 35$ in all dimensions and in different situations; filtered (boxes 1–4) and smoothed ($T = 50$; boxes 5–8) versions

situation	filter			smoother		
	Kalman	rLS.IO	rLS.AO	Kalman	rLS.IO	rLS.AO
id	0.034	0.047	1.725	0.011	0.013	1.514
AO	1995.207	7676.445	5.661	5520.957	32270.199	5.262
IO	175.484	3.553	7515.495	116.626	86.502	7513.354

Table 2: empirical MSEs in Model (SimB) at $t = 35$ for $T = 50$

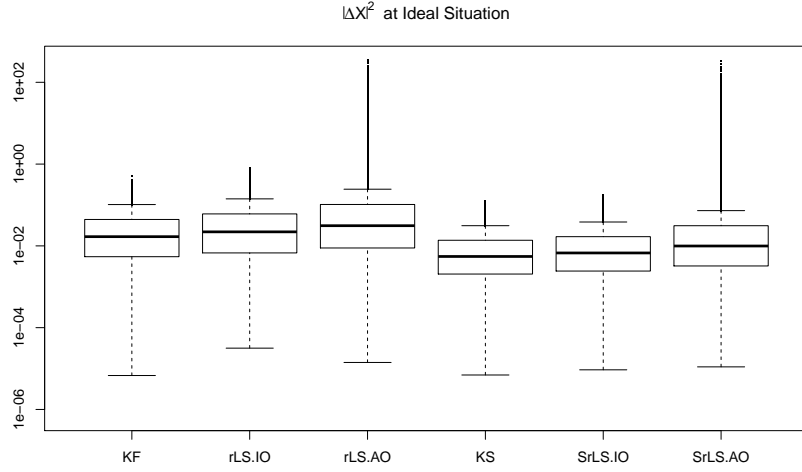


Figure 9: $\|\Delta X_t\|^2$ in Model (SimB) at $t = 35$ for $T = 50$ in ideal situation; filtered (boxes 1–3) and smoothed ($T = 50$; boxes 4–6) versions

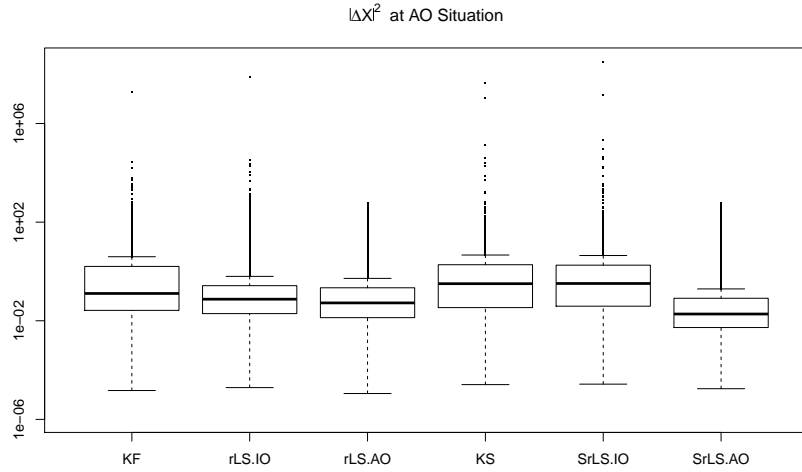


Figure 10: $\|\Delta X_t\|^2$ in Model (SimB) at $t = 35$ for $T = 50$ in AO situation; filtered (boxes 1–3) and smoothed ($T = 50$; boxes 4–6) versions

also can cover with a wider variety of outlier situations. These stylized situations also helped to identify certain flaws of the procedures.

In particular the new rLS.AO smoother and the rLS.IO filter seem to be recommendable already as they are. As we have seen in the evaluations, further research is needed though to improve the IO-robust smoother which could not convince so far.

As to the theoretical contributions, for the tracking problem we have thoroughly settled the case of non invertible observation matrices, i.e.; the situation when certain directions of the

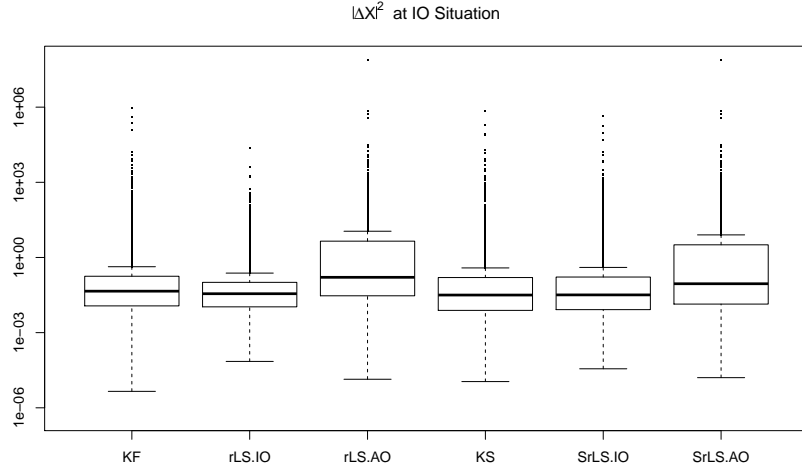


Figure 11: $\|\Delta X_t\|^2$ in Model (SimB) at $t = 35$ for $T = 50$ in IO situation; filtered (boxes 1–3) and smoothed ($T = 50$; boxes 4–6) versions

state are not visible. It turned out that the taken passage to observation space in order to define the rLS.IO involves no extra costs in terms of rank defects compared to the classical Kalman filter. In order to have a well-posed problem though, as demonstrated in Section 4, we need to pass to a semi-norm which ignores directions, no filter can see. With this modified criterion, we can establish our IO robust filter as approximately optimal, where approximately means that it would be optimal if the ideal conditional expectation were linear. Now the conditional expectation is not exactly linear, but, as shown empirically in specific examples, it is close to linear—at least in a central region.

Outlook In reality “pure” IO or AO situations hardly occur. Hence we think it is of high importance to thoroughly study hybrid versions of our filters (and/or smoothers) combining the two types of filters (IO and AO) we have discussed so far, to use them in mixed situations. So far we have only checked a heuristics based on the sequence of the normed observation residuals $\|\Delta Y_t\|$ in a rolling window. In situation where IOs and AOs are well separated this already works decently, but the procedure easily gets confused once in a window we have both IOs and AOs.

In non-linear state space models, the unscented Kalman filter, compare Wan and van der Merwe (2002) deserves a robustification, which could easily take up the ideas of this paper.

Finally, robust filters and smoothers are a key ingredient of the EM algorithm (Shumway and Stoffer, 1982) to estimate unknown (hyper-) parameters with interesting application to the fitting of stochastic differential equations to financial data.

Acknowledgements

The authors thank two anonymous referees for their valuable and helpful comments. Financial support from VW foundation in the framework of project Robust Risk Estimation for D. Pupashenko is gratefully acknowledged.

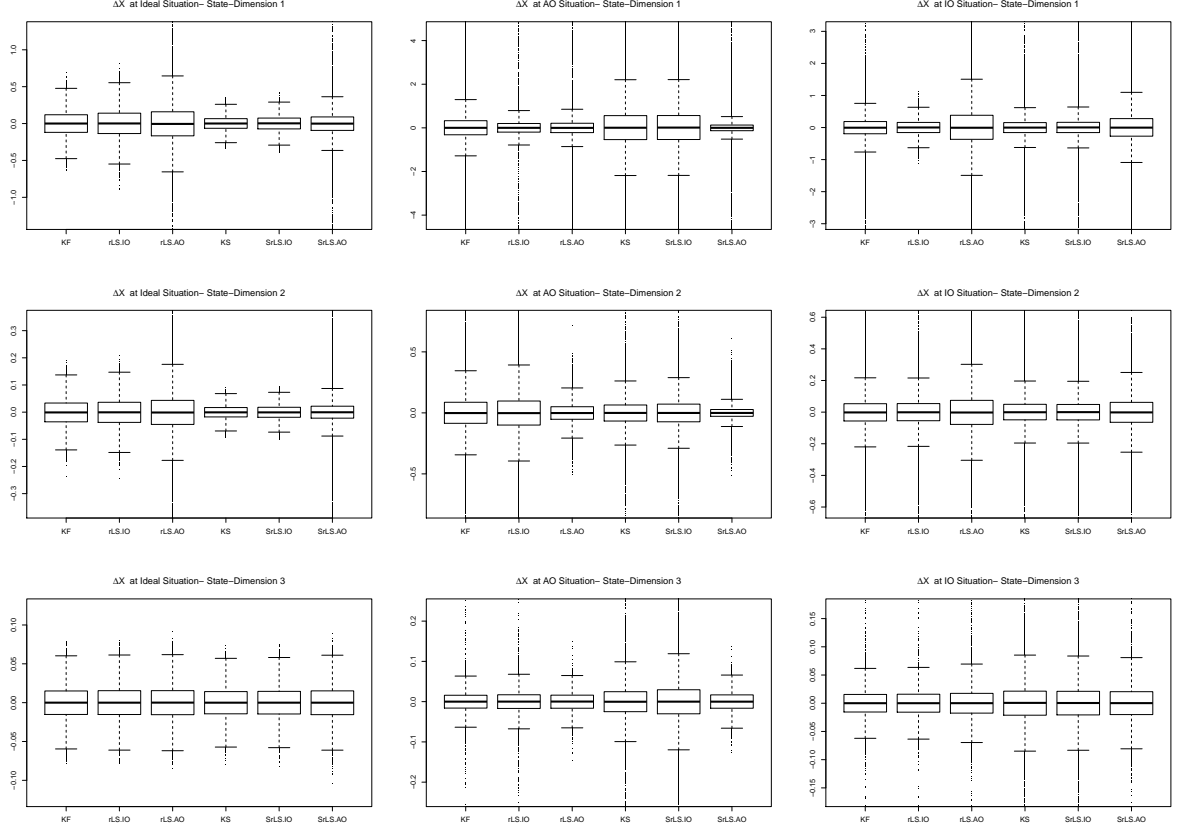


Figure 12: ΔX_t in Model (SimB) at $t = 35$ in all dimensions and in different situations; filtered (boxes 1–3) and smoothed ($T = 50$; boxes 4–6) versions

A Appendix

A.1 Optimality of the classical Kalman filter

Optimality of the classical Kalman filter among all linear filters in L_2 -sense and, under normality of the error and innovation distributions, among all measurable filters is a well-known fact, compare, e.g. Anderson and Moore (1990, Sec. 5.2). As we will need some of the arguments later, let us complement this fact by some generalization to arbitrary norms generated by a quadratic form and by a thorough treatment of the case of singularities in the covariances arising in the definition of the Kalman gain from (3.5). To do so, we take the orthogonal decomposition of the Hilbert into closed linear subspaces as $\text{lin}(Y_{1:(t-1)}) \oplus \text{lin}(\Delta Y_t)$ as granted and for $X = X_t - X_{t|t-1}$ and $Y = \Delta Y_t$ as given in (3.5), derive \hat{K}_t . To this end, for any matrix A let us denote by A^- the generalized inverse of A with the defining properties

$$A^- A A^- = A^-, \quad A A^- A = A, \quad A^- A = (A^- A)^\tau, \quad A A^- = (A A^-)^\tau \quad (\text{A.1})$$

and for D a positive semi-definite symmetric matrix in $\mathbb{R}^{p \times p}$ and for $x \in \mathbb{R}^p$ define the semi-norm generated by D as $\|x\|_D^2 := x^\tau D x$.

Lemma A.1. *Let $p, q \in \mathbb{N}$, P some probability and $X \in L_2^p(P)$, $Y \in L_2^q(P)$, $\mathbb{E} X = 0$, $\mathbb{E} Y = 0$, where for some $Z \in \mathbb{R}^{q \times p}$, and some $\varepsilon \in L_2^q(P)$ independent of X , $Y = ZX + \varepsilon$. Let D a positive*

semi-definite symmetric matrix in $\mathbb{R}^{p \times p}$. Then

$$\hat{K} = \text{Cov}(X, Y) \text{Cov}(Y)^{-} \quad (\text{A.2})$$

solves

$$\mathbb{E} \|X - KY\|_D^2 = \min!, \quad K \in \mathbb{R}^{p \times q} \quad (\text{A.3})$$

\hat{K} is unique up to addition of some $A \in \mathbb{R}^{p \times q}$ such that $A \text{Cov} Y = 0$ and some $B \in \mathbb{R}^{p \times q}$ such that $DB = 0$. If $\hat{K} = DD^{-}\hat{K}$, \hat{K} has smallest Frobenius norm among all solutions K to (A.3).

Proof. Denote $L_2^p(P, D)$ the Hilbert space generated by all \mathbb{R}^p valued random variables U such that $\mathbb{E}_P \|U\|_D^2 < \infty$ —after a passage to equivalence classes of all random variables U, U' such that $\mathbb{E}_P \|U - U'\|_D^2 = 0$. Let $S = \text{Cov}(X)$ and $V = \text{Cov}(\varepsilon)$. Then $\text{Cov}(X, Y) = SZ^T$ and $\text{Cov}(Y) = ZSZ^T + V$. Denote the approximation space $\{KY \mid K \in \mathbb{R}^{p \times q}\} \subset L_2^p(P, D)$ by \mathcal{K} . \mathcal{K} is a closed linear subspace of $L_2^p(P, D)$, hence by Rudin (1974, Thm. 4.10) there exists a unique minimizer $\hat{X} = \hat{K}Y \in \mathcal{K}$ to problem (A.3). It is characterized by

$$\mathbb{E}(X - \hat{K}Y)^T D^{-} KY = 0, \quad \forall K \in \mathbb{R}^{p \times q} \quad (\text{A.4})$$

Plugging in $K = De_i \tilde{e}_j^T$, $\{e_i\}$, $\{\tilde{e}_j\}$ canonical bases of \mathbb{R}^p , \mathbb{R}^q , respectively, we see that (A.4) is equivalent to

$$\mathbb{E} \pi(X - \hat{K}Y)Y^T = 0 \iff \pi K \text{Cov}(Y) = \pi \text{Cov}(X, Y) \quad (\text{A.5})$$

where $\pi = D^{-}D$ is the orthogonal projector onto the column space of D . But $y \in \mathbb{R}^q$ can only lie in $\ker \text{Cov}(Y)$ if $y \in \ker \text{Cov}(X, Y)$. Hence indeed $\hat{K} \text{Cov}(Y) = \text{Cov}(X, Y)$, and the uniqueness assertion is obvious. We write π_D and π_C for the orthogonal projectors to the column spaces of D and $\text{Cov}(Y)$, respectively and $\bar{\pi}_C = \mathbb{I}_q - \pi_C$, $\bar{\pi}_D = \mathbb{I}_p - \pi_D$ for the corresponding complementary projectors. Then we see that $\hat{K} = \hat{K}\pi_C$ and for any $A \in \mathbb{R}^{p \times q}$ with $A \text{Cov} Y = 0$ we have $A = A\bar{\pi}_C$ and for any $B \in \mathbb{R}^{p \times q}$ with $DB = 0$ we have $B = \bar{\pi}_D B$; hence

$$\begin{aligned} \|\hat{K} + A + B\|^2 &= \text{tr} \hat{K}^T \hat{K} + 2 \text{tr} A^T \hat{K} + 2 \text{tr} B^T \hat{K} + \text{tr}(A + B)^T (A + B) \\ &= \|\hat{K}\|^2 + 2 \text{tr} \hat{K} \pi_C \bar{\pi}_C A^T + 2 \text{tr} \hat{K} \pi_D \bar{\pi}_D B^T + \|A + B\|^2 \\ &= \|\hat{K}\|^2 + \|A + B\|^2 \geq \|\hat{K}\|^2 \quad \text{with equality iff } A + B = 0. \end{aligned}$$

□

A.2 Sketch of the optimality of the rLS.AO

(One-Step)-optimality of the rLS The rLS filter is optimally-robust in some sense: To see this, in a first step we essentially boil down our SSM to (3.9), i.e., we have an unobservable but interesting state $X \sim P^X(dx)$, where for technical reasons we assume that in the ideal model $\mathbb{E}|X|^2 < \infty$. Instead of X , for some $Z \in \mathbb{R}^{q \times p}$, we rather observe the sum $Y = ZX + \varepsilon$ of X and a stochastically independent error ε . As (wide-sense) AO model, we consider the SO outlier of (2.5), (2.6). The corresponding neighborhood is defined as

$$\mathcal{U}^{\text{SO}}(r) = \bigcup_{0 \leq s \leq r} \left\{ \mathcal{L}(X, Y^{\text{re}}) \mid Y^{\text{re}} \text{ acc. to (2.5) and (2.6) with radius } s \right\} \quad (\text{A.6})$$

In this setting we may formulate two typical robust optimization problems, i.e., a minimax formulation, and, in the spirit of Hampel (1968, Lemma 5), a formulation where robustness enters as side condition on the bias to be fulfilled on the whole neighborhood

$$[\text{Minmax-SO}] \quad \max_{\mathcal{U}} \mathbb{E}_{\text{re}} |X - f(Y^{\text{re}})|^2 = \min_f! \quad (\text{A.7})$$

$$[\text{Lemma-5}] \quad \mathbb{E}_{\text{id}} |X - f(Y^{\text{id}})|^2 = \min_f! \quad \text{s.t. } \sup_{\mathcal{U}} |\mathbb{E}_{\text{re}} f(Y^{\text{re}}) - \mathbb{E} X| \leq b \quad (\text{A.8})$$

Then one can show that setting $D(y) = \mathbb{E}_{\text{id}}[X|Y = y] - \mathbb{E} X$, the solution to both problems is $\hat{f}(y) = \mathbb{E} X + H_\rho(D(y))$ (with $b = \rho/r$ in Problem (A.8)), and that this is just the (one-step) rLS, once $\mathbb{E}_{\text{id}}[X|Y]$ is linear in Y . A proof to this assertion is given in Ruckdeschel (2010a, Thm. 3.2).

Remark A.2. (a) As mentioned in Section 3.2, Cipra and Hanzak (2011) show an optimality similar to the one for Problem (A.8), and hence, non-surprisingly come up with a similar procedure.

(b) The ACM filter by Masreliez and Martin (1977), an early competitor to the rLS, by analogy applies Huber (1964)’s minimax variance result to the “random location parameter X ” setting of (3.9). They come up with redescenders as filter f . Hence the ACM filter is not so much vulnerable in the extreme tails but rather where the corresponding ψ function takes its maximum in absolute value. Care has to be taken, as such “inliers” producing the least favorable situation for the ACM are much harder to detect on naïve data inspection, in particular in higher dimensions.

(c) For exact SO-optimality of the rLS-filter, linearity of the ideal conditional expectation is crucial. However, one can show that $E_{\text{id}}[\Delta X|\Delta Y]$ is linear iff ΔX is normal, but, having used the rLS-filter in the ΔX -past, normality cannot hold, see Ruckdeschel (2010a, Prop.’s 3.4, 3.6).

(d) Although rLS fails to be SO-optimal for $t > 1$, it does perform quite well at both simulations and real data. To some extent this can be explained by passing to a certain extension of the original SO-neighborhoods. For details see (Ruckdeschel, 2010a, Thm. 3.10, Prop. 3.11).

A.3 Optimality of the rLS.IO

This section discusses (one-step) optimality of the rLS.IO in some detail. We omit time indices and write Σ for $\Sigma_{t|t-1}$. To start, let us again look at the boiled down model (3.9) where we interchange the rôle of ε and X , and note that $X - f(Y) = \varepsilon - g(Y)$ for $f(Y) = Y - g(Y)$. Hence in this simple model, the optimal reconstruction of a corrupted X assuming that ε is still from the ideal distribution is just $Y - g(Y)$, $g(Y)$ the optimal reconstruction of ε in the same situation.

In notation, let us write $\text{oP}(a|b)$ for the best linear reconstruction of a by means of b , i.e., the orthogonal projection of a onto the closed linear space generated by b .

Assuming linear conditional expectations and mutatis mutandis in Ruckdeschel (2010a, Thm. 3.2), the optimally-robust reconstruction of ε given Y in the sense of Problems (A.7), (A.8) is just $H_b(\text{oP}(\varepsilon|Y))$ —with the same caveats as to the optimality for larger time indices as in Remark A.2. But again, $\text{oP}(\varepsilon|Y) = \text{oP}(Y - X|Y) = Y - \text{oP}(X|Y)$ so the IO-optimal procedure f_{IO} is

$$f_{\text{IO}}(Y) = Y - H_b(Y - \text{oP}(X|Y)) \quad (\text{A.9})$$

Details as to the translation of the contamination neighborhoods and exact formulations of the optimality results are given in Ruckdeschel (2010a).

The general setup with some arbitrary $Z \in \mathbb{R}^{q \times p}$, where Z in general is not invertible, and moreover, even $Z^\tau \Sigma Z$ may be singular, is not trivial, though. For instance, our preceding argument so far only covers reconstruction of ZX , but at this stage it is not obvious how to optimally derive a reconstruction of X from this. In particular, in this general case, there are directions which our (robustified) reconstruction cannot see—at least all directions in $\ker Z$. So an unbounded criterion like MSE would play havoc once unbounded contamination happens in these directions. So in this context, the best we can do is optimally reconstructing ZX on the whole neighborhood generated by outliers in X and then, in a second step, for this best reconstruction of ZX , find the best back-transform to X in the ideal model setting. The question is how much we loose by this. To this end, note that

$$\text{oP}(\varepsilon|Y) = \text{oP}(Y - ZX|Y) = Y - Z \text{oP}(X|Y) = (\mathbb{I}_q - ZK)Y \quad (\text{A.10})$$

For Z^Σ from (3.12), we introduce the orthogonal projector onto the column space of $Z^\tau \Sigma Z$ and its orthogonal complement as

$$\pi_{Z,\Sigma} = ZZ^\Sigma, \quad \bar{\pi}_{Z,\Sigma} := \mathbb{I}_q - \pi_{Z,\Sigma} \quad (\text{A.11})$$

Then we have the following Lemma:

Lemma A.3. (a) For any positive definite D , Z^Σ from (3.12) solves

$$E_{\text{id}} \|X - A \text{oP}(ZX|Y)\|_D^2 = \min!, \quad A \in \mathbb{R}^{p \times q} \quad (\text{A.12})$$

(b) $\Sigma Z^\tau \bar{\pi}_{Z,\Sigma} = 0$; in particular, no matter of the rank of Z or $\pi_{Z,\Sigma}$, with $K = \Sigma Z^\tau C^-$,

$$Z^\Sigma Z K = K \quad (\text{A.13})$$

Proof. (a) As in Lemma A.1, we see that

$$\hat{A} = \Sigma Z^\tau K^\tau Z^\tau (Z K \text{Cov}(Y) K^\tau Z^\tau)^- \quad (\text{A.14})$$

Abbreviating $Z \Sigma Z^\tau$ by B and $\text{Cov}(Y)$ by C , this gives $\hat{A} = \Sigma Z^\tau C^- B (B C^- B)^-$, and with $\Sigma_{.5}$ the symmetric root of Σ , and with $G = \Sigma_{.5} Z^\tau$, this becomes $\hat{A} = \Sigma_{.5} G C^- G^\tau G (G^\tau G C^- G^\tau G)^-$. Next we pass to the singular value decomposition of $G = U S W^\tau$, with U, W corresponding orthogonal matrices in $\mathbb{R}^{p \times p}$ and $\mathbb{R}^{q \times q}$, respectively, and $S \in \mathbb{R}^{p \times q}$ a matrix with the singular values on the “diagonal entries” $S_{i,i}$, $i = 1, \dots, \min(p, q)$ and $S_{i,j} = 0$, $i \neq j$; furthermore, $S_{i,i} > 0$ for $i = 1, \dots, d$, $d \leq \min(p, q)$ and 0 else. Using $(aba^\tau)^- = (a^\tau)^{-1} b^- a^{-1}$ for a invertible and setting $T = S^\tau S$, we obtain

$$\hat{A} = \Sigma_{.5} U S W^\tau C^- W T W^\tau (W T W^\tau C^- W T W^\tau)^- = \Sigma_{.5} U S W^\tau C^- W T (T W^\tau C^- W T)^- W^\tau$$

As the expressions of the symmetric matrices $W^\tau C^- W$ are surrounded by S (resp. T)-terms, we may replace them with a matrix $R \in \mathbb{R}^{q \times q}$ with only entries in the upper $d \times d$ block, i.e., $\hat{A} = \Sigma_{.5} U S R T (T R T)^- W^\tau$ and as R now is compatible with S and T ,

$$\hat{A} = \Sigma_{.5} U S R 1_d R^- T^- W^\tau = \Sigma_{.5} U S R R^- T^- W^\tau, \quad \text{for } 1_d = T T^-$$

Now, as $C = W T W^\tau + V$, $W^\tau C^- W = (T + W^\tau V W)^-$, in particular the upper $d \times d$ block R_d of $R = 1_d (T + W^\tau V W)^- 1_d$ is invertible and

$$\hat{A} = \Sigma_{.5} U S T^- W^\tau (= \Sigma_{.5} U S^- W^\tau) = \Sigma_{.5} U S W^\tau W T^- W^\tau = \Sigma Z^\tau B^- = Z^\Sigma$$

(b) We start by noting that $\Sigma Z^\tau \bar{\pi}_{Z,\Sigma} = \Sigma_{.5} U S W^\tau W (\mathbb{I}_q - 1_d) W^\tau = 0$. For (A.13), we write $K = \Sigma Z^\tau (\pi_{Z,\Sigma} + \bar{\pi}_{Z,\Sigma}) C^- = \Sigma Z^\tau \pi_{Z,\Sigma} C^- = Z^\Sigma Z K$. □

As a consequence of assertion (b) in the preceding Lemma, we obtain

Corollary A.4. *No matter of the rank of Z or $\pi_{Z,\Sigma}$,*

$$\text{oP}(X|Y) = Z^\Sigma (Y - \text{oP}(\varepsilon|Y)) \quad (\text{A.15})$$

that is, we can exactly recover $\text{oP}(X|Y)$ from $\text{oP}(\varepsilon|Y)$, and passing over the reconstruction of ZX first does not cost us anything in efficiency compared to the direct route.

Proof. We only note that $Y - \text{oP}(\varepsilon|Y) = \text{oP}(ZX|Y) = ZKY$. □

To keep things well-defined in this setting where we have “invisible directions” in the state, we may recur to passing to a semi-norm in X -space which ignores such directions. A possible candidate for D in Lemma A.1 is

$$D^- = (Z^\Sigma Z)^\tau \Sigma^- Z^\Sigma Z \quad (\text{A.16})$$

On the one hand, as we show below, invisible directions get ignored, on the other hand, by (A.13), no direction visible for the classically optimal procedure is lost.

Proposition A.5. *Using D from (A.16) and assuming observation errors from the ideal situation, maximal MSE error for rLS.IO measured in this norm remains bounded for IO contamination. With this norm, \hat{K} is smallest possible solution to (A.3) in Frobenius norm.*

Proof. The error term $e = X - \hat{X}$ for the rLS.IO can be written as

$$e = X - Z^\Sigma(Y - H_b(Y - ZKY)) = (\mathbb{I}_p - Z^\Sigma Z)X - Z^\Sigma(\varepsilon - H_b((\mathbb{I}_q - ZK)Y))$$

As $(Z^\Sigma Z)^2 = Z^\Sigma Z$, we see that $(\mathbb{I}_p - Z^\Sigma Z)(Z^\Sigma Z) = 0$, so that in D -semi-norm, the $(\mathbb{I}_p - Z^\Sigma Z)X$ terms cancel out and we get

$$\begin{aligned} e^\tau D^- e &= [Z^\Sigma(\varepsilon - H_b(\cdot))]^\tau (Z^\Sigma Z)^\tau \Sigma^- Z^\Sigma Z [Z^\Sigma(\varepsilon - H_b(\cdot))] = \\ &= (\varepsilon - H_b(\cdot))^\tau (Z^\Sigma)^\tau (Z^\Sigma Z)^\tau \Sigma^- Z^\Sigma Z Z^\Sigma (\varepsilon - H_b(\cdot)) = \\ &= (\varepsilon - H_b(\cdot))^\tau B^- (\varepsilon - H_b(\cdot)) \leq 2\varepsilon^\tau B^- \varepsilon + 2H_b(\cdot)^\tau B^- H_b(\cdot) \end{aligned}$$

so MSE is bounded by $2\text{tr}(B^-(V + b^2\mathbb{I}_q))$. The second assertion is an immediate consequence of Lemma A.1 and Lemma A.3(b). \square

Note that changing the norm in the Y -space is not necessary for boundedness reasons, as with only ideally distributed ε , the reconstruction of ZX can be achieved such that no matter how largely ΔX is contaminated, the maximal MSE remains bounded.

References

- Anderson B.D.O. and Moore J.B. (1990): *Optimal control. Linear quadratic methods*. Prentice Hall.
- Anscombe F.J. (1960): Rejection of outliers. *Technometrics*, **2**: 123–147.
- Birmiwal K. and Papantoni-Kazakos P. (1994): Outlier resistant prediction for stationary processes. *Stat. Decis.*, **12**(4): 395–427.
- Birmiwal K. and Shen J. (1993): Optimal robust filtering. *Stat. Decis.*, **11**(2): 101–119.
- Boncellet C.G. Jr and Dickinson B.W. (1983): An approach to robust Kalman filtering. In: *Proceedings on the 22nd IEEE Conference on Decision & Control*, Vol. 1, pp. 304–305.
- (1987): An extension to the SRIF Kalman filter. *IEEE Trans. Autom. Control*, **AC-32**: 176–179.
- Cipra T. and Hanzak T. (2011): Exponential smoothing for a time series with outliers. *Kybernetika*, **47**: 165–178.
- Cipra T. and Romera R. (1991): Robust Kalman filter and its application in time series analysis. *Kybernetika*, **27**(6): 481–494.
- Donoho D. and Johnstone I. (1994): Ideal patial adaptation via wavelet shrinkage. *Biometrika*, **81**: 425–455.
- Ershov A.A. and Liptser R.S. (1978): Robust Kalman filter in discrete time. *Autom. Remote Control*, **39**: 359–367.
- Fox A.J. (1972): Outliers in time series. *J. R. Stat. Soc., Ser. B*, **34**: 350–363.
- Franke J. (1985): Minimax-robust prediction of discrete time series. *Z. Wahrscheinlichkeitstheor. Verw. Geb.*, **68**: 337–364.
- Franke J. and Poor H.V. (1984): Minimax-robust filtering and finite-length robust predictors. In: *Robust and nonlinear time series analysis. Proc. Workshop, Heidelberg/Ger. 1983*, Nr. 26 in Lect. Notes Stat. Springer.
- Fried R., Bernholt T. and Gather U. (2006): Repeated Median and Hybrid Filters. *Computational Statistics & Data Analysis*, **50**: 2313–2338.

- Fried R., Einbeck J. and Gather U. (2007): Weighted repeated median smoothing and filtering. *J. Am. Stat. Assoc.*, **102**: 1300–1308.
- Fried R. and Schettlinger K. (2010): *robfilter: Robust Time Series Filters*. R package version 2.6.1. Available on CRAN. <http://cran.r-project.org/web/packages/robfilter>.
- Gelper S., Fried R. and Croux C. (2010): Robust forecasting with exponential and holt-winters smoothing. *J. Forecasting*, **29**: 285–300.
- Genz A., Bretz F., Miwa T., Mi X., Leisch F., Scheipl F. and Hothorn T. (2011): *mvtnorm: Multivariate Normal and t Distributions*. R package version 0.9-9991. Available on CRAN. <http://CRAN.R-project.org/package=mvtnorm>.
- Hampel F.R. (1968): *Contributions to the theory of robust estimation*. Dissertation, University of California, Berkely, CA.
- Hampel F.R., Ronchetti E.M., Rousseeuw P.J. and Stahel W.A. (1986): *Robust statistics. The approach based on influence functions*. Wiley.
- Huber P.J. (1964): Robust estimation of a location parameter. *Ann. Math. Stat.*, **35**: 73–101.
- Kalman R.E. (1960): A new approach to linear filtering and prediction problems. *Journal of Basic Engineering—Transactions of the ASME*, **82**: 35–45.
- Kassam S.A. and Poor H.V. (1985): Robust techniques for signal processing: A survey. *Proc. IEEE*, **73**(3): 433–481.
- Künsch H. (2001): State space models and Hidden Markov Models. In: Barndorff-Nielsen O.E., Cox D.R. and Klüppelberg C. (Eds.) *Complex Stochastic Systems*, pp. 109–173. Chapman and Hall.
- Martin R.D. and Yohai V.J. (1986): Influence functionals for time series (with discussion). *Ann. Stat.*, **14**: 781–818.
- Masreliez C.J. and Martin R. (1977): Robust Bayesian estimation for the linear model and robustifying the Kalman filter. *IEEE Trans. Autom. Control*, **AC-22**: 361–371.
- R Development Core Team (2012): *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. <http://www.R-project.org>.
- Rieder H., Kohl M. and Ruckdeschel P. (2008): The costs of not knowing the radius. *Statistical Methods and Applications*, **17**(1): 13–40.
- Ruckdeschel P. (2000): Robust Kalman filtering. In: Härdle W., Hlávka Z. and Klink S. (Eds.) *XploRe. Application Guide.*, Chap. 18, pp. 483–516. Springer.
- (2001): *Ansätze zur Robustifizierung des Kalman-Filters*, Vol. 64. Bayreuther Mathematische Schriften, Bayreuth.
- (2010a): Optimally (distributional-)robust Kalman filtering. Available on arXiv: 1004.3393.
- (2010b): Optimally robust Kalman filtering. Techn. Report 185, Fraunhofer ITWM Kaiserslautern, Fraunhofer Platz 1, 67663 Kaiserslautern, Germany. http://www.itwm.fraunhofer.de/fileadmin/ITWM-Media/Zentral/Pdf/Berichte_ITWM/2010/bericht_185.pdf.
- Rudin W. (1974): *Real and complex analysis*. McGraw-Hill, 2nd edition.
- Schettlinger K. (2009): *Signal and variability extraction for online monitoring in intensive care*. Dissertation, TU Dortmund, Dortmund.

- Schick I.C. and Mitter S.K. (1994): Robust recursive estimation in the presence of heavy-tailed observation noise. *Ann. Stat.*, **22**(2): 1045–1080.
- Shumway R.H. and Stoffer D.S. (1982): An approach to time series smoothing and forecasting using the EM algorithm. *J. Time Ser. Anal.*, **3**: 253–264.
- Spangl B. (2008): *On Robust Spectral Density Estimation*. Dissertation, Dept. of Statistics and Probability Theory, Vienna University of Technology, Vienna.
- Stockinger N. and Dutter R. (1987): Robust time series analysis: A survey. *Kybernetika*, **23**. Supplement.
- Venables W. and Ripley B. (2002): *Modern Applied Statistics with S-Plus*. Springer, 4th edition.
- Wan E.A. and van der Merwe R. (2002): The Unscented Kalman Filter. In: Haykin S. (Ed.) *Kalman Filtering and Neural Networks*. Wiley.